

EXISTENCE, CHARACTERIZATION AND APPROXIMATION IN THE GENERALIZED MONOTONE-FOLLOWER PROBLEM

JIEXIAN LI AND GORDAN ŽITKOVIĆ

ABSTRACT. We revisit the classical monotone-follower problem and consider it in a generalized formulation. Our approach is based on a compactness substitute for nondecreasing processes, the Meyer-Zheng weak convergence, and the maximum principle of Pontryagin. It establishes existence under weak conditions, produces general approximation results and further elucidates the celebrated connection between singular stochastic control and stopping.

1. INTRODUCTION

A direct precursor to the monotone-follower problem dates back to the 1970's; the basic model originated from engineering and first appeared in the work of Bather and Chernoff [BC67]. There, it was posed in a model of a spaceship being steered towards a target with both precision and fuel consumption appearing in the performance criterion. The authors observed an unexpected connection between the control problem they studied and a Brownian optimal stopping problem based on the same ingredients; arguing quite incisively, but mostly on heuristic grounds, they demonstrated that the value function of the latter is the gradient of the value function of the former.

In 1984, Karatzas and Shreve [KS84] considered a generalized version of the Bather-Chernoff problem dubbing it the “monotone follower problem”. In the same paper, using purely probabilistic tools, they established rigorously the equivalence of the control and stopping problems under appropriate continuity and growth conditions. Some time later, Haussmann and Suo [HS95] applied relaxation and compactification methods, used the Meyer-Zheng convergence, and showed existence of the optimal control under a different set of conditions. In 2005, Bank [Ban05] constructed a fairly explicit control policy under stochastic dynamic fuel constraint in one dimension. Subsequently, Budhiraja and Ross [BR06] applied the Meyer-Zheng convergence to prove a general existence theorem, also

Date: October 14, 2016.

2010 Mathematics Subject Classification. 93E20.

Key words and phrases. maximum principle, Meyer-Zheng convergence, monotone-follower problem, optimal stochastic control, optimal stopping, singular control.

Acknowledgements: Both authors would like to thank Ioannis Karatzas and Mihai Sirbu for valuable conversations, and the anonymous referee for a simplification of the proof of Theorem 2.7 and other insightful comments. The second author, furthermore, acknowledges the support by the National Science Foundation under Grants No. DMS-0706947 (2010 - 2015), No. DMS-1107465 (2012 - 2017) and No. DMS-1516165 (2015-2018). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

under a fuel constraint. Guo and Tomecek [GT08] generalized some results of [KS84] in a different direction: they established a connection between singular control of finite variation and optimal switching.

Problem formulation. The essence of the monotone follower problem is tracking, as closely as possible, a given random process L (the *target*) by a suitably constrained control process A (the *follower*). In the original setting of [KS84], the target is a Brownian motion, the follower is required to be adapted and non-decreasing, and the “closeness” is measured by applying an appropriate functional to the state variable defined as the difference between the position of the target and the position of the follower. Our version of this problem is generalized in two directions:

(a) We allow the dynamics of both the target and the follower to be multidimensional and impose weak assumptions on the distribution of dynamics the target L . For our existence and characterization results (Theorems 2.7 and 2.12 below), we only require that L has càdlàg paths. For the approximation (Theorem A.5 below), we need L to be a Feller process (still allowing, in particular, inhomogeneities in the cost structure). Also, we consider functionals which are functions of the target and the follower, convex in the position of the follower, and not only functions of their relative positions. Finally, we relax some of the growth assumptions; in particular, we do not require superlinear growth of the cost function to obtain existence of an optimal control (as in, e.g., [Ban05], where it serves as a sufficient condition for the existence of a solution to a stochastic representation problem which, in turn, characterizes the optimizer.)

(b) Our formulation is weak (distributional), in the sense that we are only interested in the joint distribution of the follower and the target, without fixing the underlying filtered probability space and making it a part of the problem. This enables us to prove an approximation result (Theorem A.5 below) in great generality. On the other hand, as we will see below in Proposition 3.1, every weak (distributional) solution can be turned into a strong one under usually met conditions by a simple projection operation. Moreover, as far as generality is concerned, any setup where the filtration is generated by a finite number of càdlàg processes can be easily lifted to our canonical framework, allowing us to work with on a canonical (Skorokhod) space right from the start. It is worth noting that (even though we do not provide details for such an approach here) even greater generality can be achieved by considering Polish-space-valued càdlàg processes and their natural filtrations.

Our results. We treat questions of existence, approximability and characterization (via Pontryagin’s maximum principle), as well as connections with optimal stopping. These are tackled using a variety of methods, including a compactness substitute for monotone processes and the Meyer-Zheng convergence. Moreover, we posit the idea that

*the connection between control and stopping can be understood as the connection
between the monotone-follower problem and its Pontryagin maximum principle.*

The original impetus for our research was twofold:

(a) On the one hand, we wanted to understand the role played by different regularity and growth conditions imposed in the existing literature in order to establish existence of optimal controls. This lead to an existence proof (Theorem 2.7 below) under less restrictive conditions on most ingredients.

The proof is based on a convenient substitute for compactness under convexity, and not on the Meyer-Zheng topology as in some of the papers mentioned above. The beginnings of such an approach can be traced back to the fundamental result of Komlós [Sch86], while the version used in the present paper is due to Kabanov [Kab99].

(b) On other hand - perhaps more importantly - we tried to grasp a more practical issue better, namely, the approximation of the archetypically singular monotone-follower problem by a sequence of regular, absolutely continuous (even Lipschitz) control problems. To accomplish this task, the following conceptual framework was devised. First, a sequence of so-called “capped” problems where the exerted controls are constrained to be Lipschitz is posed. These *regular* problems come with increasing upper bounds on the Lipschitz constant and are expected to approach the monotone control problem both in value and in optimal controls. Being regular and well-behaved, each capped problem is expected to be solvable by the well-known classical methods; the resulting solution sequence is, then, expected to converge (in the appropriate sense) towards the solution to the original problem.

The second, larger, part of the paper can be seen as the implementation of the above steps. The major difficulty we encountered was the lack of good equicontinuity estimates on the solutions to the capped problems. To overcome it we replaced the usual weak convergence under the Skorokhod topology with the versatile Meyer-Zheng convergence. Even so, we still needed to close the gap between the limit of the values of the capped problems and the value of the original problem. For that, we characterized the optimizers (both in the capped and the original problems) via the maximum principle of Pontryagin (i.e., the “first-order” condition) and passed to a Meyer-Zheng limit there.

While ideas described in the previous paragraphs seem to be new, the research relating Pontryagin’s maximum principle to singular control problems is certainly not. Indeed, the Pontryagin’s maximum principle for singular control problems was first discussed by Cadenillas and Haussmann [CH94] already in 1994. With Brownian dynamics, convex cost, and state constraints assumed, these authors formulated the stochastic maximum principle in an integral form and gave necessary and sufficient conditions for optimality. In order to solve the approximation problem via maximum principle, however, one must go beyond their work. Even though the last 20 years have seen an explosion in activity in the general theory of BSDE and FBSDE (see e.g., [MPY94], [CM96], [MY99], [MC01], [AM03], [MZ11]), to the best of our knowledge none of the existing work seems to be able to deal directly with the singular FBSDE that the maximum principle for the monotone-follower problems yields, even in the Brownian case. Our route, via approximation and simultaneous consideration of the related (capped) control problems, can be interpreted as a variational approach to a class of singular FBSDE and may, perhaps, be of use in other situations, as well. For example, a combination (see Corollary 2.14 below) of our existence and characterization results, i.e., Theorems 2.7 and 2.12, guarantees existence of solutions of such FBSDE under weak, monotonicity- and exponential-growth-type assumptions on the nonlinearities.

The approximation result (Theorem 2.21 below) serves as a pleasant justification of singular controls as a conceptual limit of absolutely continuous controls. Moreover, together with the related maximum-principle characterization of the optimal controls in the original problem, it leads us to

view the celebrated connection between stopping and control in a new light. Indeed, once such a characterization is formulated, it is a simple observation that it can be re-interpreted as an optimal stopping problem, which turns out to be precisely the optimal stopping problem identified by Bather and Chernoff and rigorously studied by Karatzas and Shreve.

Organization of the paper. After this Introduction, Section 2. contains the formulation of the problem, a description of the probabilistic setup it is defined on, and main results. Section 3. is devoted to proofs. At the end, a short compendium of the most important well-known results - including the tightness criteria - on the Meyer-Zheng topology is given in Appendix A.

2. THE PROBLEM AND THE MAIN RESULTS

2.1. Notational conventions and the canonical setup. For $N \in \mathbb{N}$, let \mathcal{D}^N denote the Skorokhod space, i.e., the measurable space of all \mathbb{R}^N -valued càdlàg functions on $[0, T]$, equipped with the σ -algebra generated by the coordinate maps. Since the same σ -algebra appears as the Borel σ -algebra generated by the Skorokhod topology, as well as by most of the other popular topologies on \mathcal{D}^N , we call it simply the Borel σ -algebra. The set of all probability measures on the Borel σ -algebra of \mathcal{D}^N is denoted by \mathfrak{P}^N . The probabilistic notation $\mathbb{E}^{\mathbb{P}}[\cdot]$ is used to denote the integration with respect to a probability measure in \mathfrak{P}^N .

The components of the coordinate process X on \mathcal{D}^N are generally denoted by X^1, \dots, X^N . Given a subset $(X^{i_1}, \dots, X^{i_K})$, with $K \leq N$, of the components of X , we denote by $\pi_{X^{i_1}, \dots, X^{i_K}}$ the projection map $\mathcal{D}^N \rightarrow \mathcal{D}^K$. For $\mathbb{P} \in \mathfrak{P}^N$, $\pi_{X^{i_1}, \dots, X^{i_K}}$ induces a probability measure on \mathcal{D}^K , which we call the $(X^{i_1}, \dots, X^{i_K})$ -marginal of \mathbb{P} and denote simply by $\mathbb{P}_{X^{i_1}, \dots, X^{i_K}}$.

Often, we group sets of variables into single-named vector-valued components to increase readability. The dimensionality of these components will always be clear from the context, with the definition of the marginal extending naturally. To make it easier for the reader, we often employ the notation of the form $\mathcal{D}^{d+k}(L, A)$ or $\mathfrak{P}^{d+k}(L, A)$ to signal the fact that the first d coordinates are collectively denoted by L and the remaining k by A . In the same spirit, we consider (raw) filtrations of the form $\mathbb{F}^Y = \{\mathcal{F}_t^Y\}_{t \in [0, T]}$, $\mathcal{F}_t^Y = \sigma(Y_s^1, \dots, Y_s^K; s \leq t)$, $t \in [0, T]$, on \mathcal{D}^N , with Y denoting some (or all) components of X . The notation for their right-continuous enlargements is $\bar{\mathbb{F}}_+^Y = \{\bar{\mathcal{F}}_{t+}^Y\}_{t \in [0, T]}$, where $\bar{\mathcal{F}}_{t+}^Y = \cap_{s>t} \mathcal{F}_s^Y$. Unless explicitly stated otherwise, the usual conditions of right-continuity and completeness *are not* assumed. When the filtration is, indeed, completed, and the measure \mathbb{P} under which the filtration is completed is clear from the context, we add a bar above \mathbb{F} (as in $\bar{\mathbb{F}}_+^Y$, e.g.).

Some of the components of the coordinate process will naturally come with further constraints, most often in the form of monotonicity: the subset \mathcal{D}_\uparrow^N of \mathcal{D}^N denotes the class of (component-wise) nondecreasing paths A with $A_0 \geq 0$ (this is natural in our context because we will think of all functions as taking the value 0 on $(-\infty, 0)$). If monotonicity is required only for a subset of components, the suggestive notation $\mathcal{D}_{\uparrow}^{K_1+K_2}$ is used. The intended meaning is that only the last K_2 components are assumed to be nondecreasing. Similarly, if the monotonicity requirement is replaced by that of finite variation, the resulting family is denoted by \mathcal{D}_{fv}^K (unlike in the case of \mathcal{D}_\uparrow^K ,

no nonnegativity requirement on A_0 is imposed for \mathcal{D}_{fv}^K). Analogous notation will be used for sets of probability measures, as well.

For $A \in \mathcal{D}_{fv}^1$ and a measurable (sufficiently integrable) function $f : [0, T] \rightarrow \mathbb{R}$, we use the appropriately-adjusted version of the Stieltjes integral. Namely, we define

$$\int_{[0, T]} f(t) dA_t := f(0)A_0 + \int_0^T f(t) dA_t,$$

where the integral on the right is the standard Lebesgue-Stieltjes integral on $(0, T]$, of f with respect to A . This corresponds to the interpretation of the process A as having a jump of size A_0 just prior to time 0. This way, we can incorporate an initial jump in the process A while staying in the standard càdlàg framework; the price we are comfortable with paying is that the implicit value $A_{0-} = 0$ has to be fixed. For multidimensional integrators and integrands, the same conventions will be used, with the usual interpretation of the multivariate integral as the sum of the component-wise integrals.

2.2. The monotone-follower problem. Given $d, k \in \mathbb{N}$, we consider the path space $\mathcal{D}_{\cdot, \uparrow}^{d+k}(L, A)$, where L plays the role of the target and A the (controlled) monotone follower. As mentioned above, the natural, raw, σ -algebras generated by the processes L and A are denoted by $\mathbb{F}^L = \{\mathcal{F}_t^L\}_{t \in [0, T]}$ and $\mathbb{F}^A = \{\mathcal{F}_t^A\}_{t \in [0, T]}$, respectively. A central object in the problem's setup is the probability measure \mathbb{P}_0 on \mathcal{D}^d which we interpret as the law of the dynamics of the target. No additional assumptions are placed on it at this point, but for some of our results to hold, we will need to require more structure later. On the other hand, all our results go through if L is assumed to take values in a Hausdorff locally-compact topological space with countable base instead of \mathbb{R}^d , but we keep everything Euclidean for simplicity.

In the spirit of our weak approach, we control the follower by choosing its joint distribution with the target L , in a suitably defined admissibility class. In the definition below, the condition $\mathbb{P}_L = \mathbb{P}_0$ ensures that L has the prescribed marginal distribution, while the conditional-independence requirement imposes a form of non-anticipativity on the control:

Definition 2.1 (Admissible controls). A probability $\mathbb{P} \in \mathfrak{P}_{\cdot, \uparrow}^{d+k}(L, A)$ is called **admissible**, denoted by $\mathbb{P} \in \mathcal{A}$, if

- (1) $\mathbb{P}_L = \mathbb{P}_0$, and
- (2) for each $t \geq 0$, conditionally on \mathcal{F}_{t+}^L , the σ -algebras \mathcal{F}_t^A and \mathcal{F}_T^L are \mathbb{P} -independent.

If, additionally, $\mathcal{F}_t^A \subseteq \mathcal{F}_{t+}^L$, for all $t \in [0, T]$, up to \mathbb{P} -negligible sets, we say that \mathbb{P} is **strongly admissible**.

Remark 2.2. The condition (2) in Definition 2.1 above can be thought of as a non-anticipativity constraint where additional, L -independent, randomization is allowed; it is a version of the so-called *hypothesis* (\mathcal{H}) of Brémaud and Yor (see [BY78]). We point out that the choice of the right-continuous augmentation \mathcal{F}_{t+}^L is crucial for our results to hold (see Example 2.9 below), but also that it reverts to the usual hypothesis (\mathcal{H}) as soon as a version of the Blumenthal's 0-1 law holds for L .

The quality of the tracking job is measured by a nonnegative convex cost functional:

Definition 2.3 (Cost functionals). A map $C : \mathcal{D}_{\cdot, \uparrow}^{d+k}(L, A) \rightarrow [0, \infty]$, is called a **cost functional** if there exist measurable functions

$$f : [0, T] \rightarrow [0, \infty)^k, g : \mathbb{R}^d \times [0, \infty)^k \rightarrow [0, \infty) \text{ and } h : \mathbb{R}^d \times [0, \infty)^k \rightarrow [0, \infty),$$

such that f is continuous, $h(l, \cdot)$, and $g(l, \cdot)$ are convex on $[0, \infty)^k$, for each $l \in \mathbb{R}^d$, and

$$C(L, A) = \int_{[0, T]} f(t) dA_t + \int_0^T h(L_t, A_t) dt + g(L_T, A_T).$$

Remark 2.4. The role of the process L in the cost functional C above is two-fold. Some of its components play the role of the target to be tracked, while the others allow the functions h and g to depend on time or on the randomness from the environment. We enforce this interpretation in the sequel by making as few assumptions on L as possible, in particular about its relation to A . See Remark, 2.22, (2), as well.

Definition 2.5 (Cost associated with a control). Given a cost functional C and an admissible probability $\mathbb{P} \in \mathcal{A}$, the **(expected) cost** $J(\mathbb{P})$ of \mathbb{P} is given by

$$J(\mathbb{P}) = \mathbb{E}^{\mathbb{P}}[C(L, A)] \in [0, \infty],$$

where L and A denote the components of dimensions d and k , respectively, in $\mathcal{D}_{\cdot, \uparrow}^{d+k}$.

Definition 2.6 (Value and solution concepts). The **value** of the monotone-follower problem is given by

$$V = \inf_{\mathbb{P} \in \mathcal{A}} J(\mathbb{P}).$$

A probability measure $\hat{\mathbb{P}} \in \mathcal{A}$ is said to be a **weak solution** to the monotone-follower problem if $J(\hat{\mathbb{P}}) < \infty$ and $V = J(\hat{\mathbb{P}})$. If such $\hat{\mathbb{P}}$ is strongly admissible, we say that the solution is **strong**. For $\varepsilon > 0$, a (weak or strong) ε -**optimal solution** is a (weakly or strongly) admissible \mathbb{P} with $J(\mathbb{P}) < V + \varepsilon$.

2.3. An existence result. Our first result establishes existence in the monotone-follower problem (Definition 2.6) under weak conditions. Here, and in the sequel, $|\cdot|$ denotes the Euclidean norm on \mathbb{R}^k .

Theorem 2.7 (Existence under linear coercivity). *Suppose that the cost function C is linearly coercive, i.e., that there exist constants $\kappa, K > 0$ such that*

$$(2.1) \quad \mathbb{E}^{\mathbb{P}}[C(L, A)] \geq \kappa \mathbb{E}^{\mathbb{P}}[|A_T|], \text{ for all } \mathbb{P} \in \mathcal{A} \text{ with } \mathbb{E}^{\mathbb{P}}[|A_T|] \geq K.$$

Then the monotone-follower problem admits a strong solution whenever its value is finite.

Remark 2.8. The reader will immediately notice that the linear coercivity condition (2.1) is a fairly weak requirement, guaranteed by either strict positivity of f , or uniform (over l) boundedness from below of the function g by a strictly increasing linear function in a , for large a . Small modifications of our results can be made to deal with the case $g = 0$, when similar, linear, coercivity is asked of h . Similarly, one can relax (2.1) even further by passing to an equivalent probability measure on the right-hand side. We leave details to the reader who comes across a situation in which such an extension is needed.

The following two examples show that neither one of the two major conditions - linear coercivity of (2.1) in Theorem 2.7, or the use of the right-continuous augmentation \mathcal{F}_{t+}^L in the definition of admissibility (Definition 2.1) - can be significantly relaxed:

Example 2.9 (Necessity of assumptions). As for the coercivity assumption (2.1), a trivial example can be constructed with $T = 1$, $f = 0$, $h = 0$, $k = d = 1$, $g(l, a) = e^{-a}$ and an arbitrary \mathbb{P}_0 . The value of the problem is clearly 0, but no minimizer exists. Linear coercivity clearly fails, too.

In order to argue that the right-continuous augmentation \mathbb{F}_+^L in the Definition 2.1 is necessary, we take $T = 1$ and assume that the dynamics of the target satisfies

$$L_t = tL_1 \text{ for } t \in [0, 1], \mathbb{P}_0\text{-a.s., with } \mathbb{P}_0[L_1 = 1] = \mathbb{P}_0[L_1 = 0] = \frac{1}{2},$$

and that the cost functional is given by

$$C(L, A) = \int_{[0,1]} \left(\frac{1}{2} + t\right) dA_t + \int_0^1 |L_t - A_t| dt.$$

Let $\mathbb{P}^* \in \mathfrak{P}_{\cdot, \uparrow}^{d+k}(L, A)$ be such that $\mathbb{P}_L^* = \mathbb{P}_0$ and $A_t^* = \frac{1}{4}L_1$, for all $t \in [0, 1]$, \mathbb{P}^* -a.s. Since the admissibility requires that $\sigma(A_0)$ and $\sigma(L_T)$ be independent, \mathbb{P}^* is not admissible. It does have the property that

$$(2.2) \quad J(\mathbb{P}^*) \leq J(\mathbb{P}), \text{ for each } \mathbb{P} \in \mathfrak{P}_{\cdot, \uparrow}^{d+k} \text{ with } \mathbb{P}_L = \mathbb{P}_0.$$

Indeed, one can check that

$$C(0, 0) \leq C(0, \alpha) \text{ and } C(\iota, \frac{1}{4}\iota) \leq C(\iota, \alpha) \text{ for all } \alpha \in \mathcal{D}_{\uparrow}^1,$$

where ι and 0 denote the identity and the constant 0 function on $[0, 1]$, respectively. Moreover, the inequality in (2.2) is an equality if and only if $\mathbb{P} = \mathbb{P}^*$. Thus, to show that no admissible minimizer exists it will be enough to find a sequence $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$ in \mathcal{A} such that $J(\mathbb{P}_n) \searrow J(\mathbb{P}^*)$. This can be achieved easily by using the \mathbb{P}_0 -laws of (A^n, L) , where

$$A_t^n = \begin{cases} \frac{1}{2}, & t < \frac{1}{n}, \\ \frac{1}{4}L_t, & t \in [\frac{1}{n}, 1], \end{cases}$$

2.4. A characterization result. Using the same ingredients as in the formulation of the monotone-follower problem, we pose a forward-backward-type stochastic equation (called the **Pontryagin FBSDE**), as a formulation of the maximum principle of Pontryagin. Whenever the Pontryagin FBSDE is involved, we automatically assume that both $a \mapsto h(l, a)$ and $a \mapsto g(l, a)$ are continuously differentiable in a on $[0, \infty)^k$ for each l , and denote their gradients (in a) by ∇h and ∇g , respectively. Any inequalities between multidimensional processes are to be understood componentwise.

Definition 2.10 (The Pontryagin FBSDE). A probability $\tilde{\mathbb{P}} \in \mathfrak{P}^{d+2k}(L, A, Y)$ is said to be a **weak solution** of the Pontryagin FBSDE if

- (1) $\tilde{\mathbb{P}}_{L, A} \in \mathcal{A}$,
- (2) $Y \geq 0$ and $\int_0^T Y_t dA_t = 0$, $\tilde{\mathbb{P}}$ -a.s.
- (3) $Y + \int_0^\cdot \nabla h(L_t, A_t) dt - f$ is an $(\mathbb{F}^{L, A, Y}, \tilde{\mathbb{P}})$ -martingale with $Y_T = f(T) + \nabla g(L_T, A_T)$, $\tilde{\mathbb{P}}$ -a.s.

Remark 2.11. Under $\tilde{\mathbb{P}}$ as above, (L, A, Y) can be interpreted as a (weak) solution to a fully-coupled stochastic forward-backward differential equation with reflection. Indeed, the forward component (L, A) feeds into the backward component Y directly (and through the terminal condition). On the other hand, the backward component affects the forward component through the reflection term in Definition 2.10, (2). The usual stochastic-representation parameter Z is hidden in our formulation (in the martingale property of Y as we do not assume the predictable-representation property in any form) and it does not feed directly into the dynamics. For that reason, it would perhaps be more appropriate to call (1)-(3) above a forward-backward stochastic equation (FBSE) instead of FBSDE; we choose to stick to the canonical nomenclature, nevertheless.

The main significance of the Pontryagin FBSDE lies in the following characterization:

Theorem 2.12 (Characterization via the Pontryagin FBSDE). *Suppose that the functions $g(l, \cdot)$ and $h(l, \cdot)$ are convex and continuously differentiable on $[0, \infty)^k$ for each $l \in \mathbb{R}^d$.*

- (1) *Suppose that there exist Borel functions $\Phi_g, \Phi_h : \mathbb{R}^d \rightarrow [0, \infty)$ and a constant $M \geq 0$, such that*

$$\int_0^T \Phi_h(L_t) dt + \Phi_g(L_T) \in \mathbb{L}^1(\mathbb{P}_L),$$

and, for $\varphi \in \{g, h\}$,

$$|\nabla \varphi(l, a)| \leq \Phi_\varphi(l) + M\varphi(l, a), \text{ for all } (l, a) \in \mathbb{R}^k \times [0, \infty).$$

Then each solution $\hat{\mathbb{P}}$ of the monotone follower problem is an (L, A) -marginal of some solution $\tilde{\mathbb{P}}$ to the Pontryagin FBSDE.

- (2) *If the Pontryagin FBSDE admits a solution $\tilde{\mathbb{P}}$, then its marginal $\tilde{\mathbb{P}}_{L,A}$ is a solution of the monotone-follower problem whenever its value is finite*

Remark 2.13.

- (1) Our Pontryagin FBSDE can be interpreted as a weakly-formulated version of (stochastic) first-order conditions. These can be found in the literature, in settings similar to ours, and in the context of singular control, e.g., in [BR01], [Ban05], or, more recently, [Ste12]).
- (2) The condition in (1) above essentially states that φ grows no faster than an exponential function, with the parameter uniformly bounded from above in l . This should be compared to virtually no growth condition needed for existence in Theorem 2.7, as well as to the polynomial growth conditions needed for the approximation result in Theorem 2.21 below.

While we will be using the Pontryagin FBSDE mostly as a tool in the proof of Theorem 2.21, we believe that the the following result, which is an immediate consequence of Theorems 2.7 and 2.12 above merits to be mentioned in its own right.

Corollary 2.14 (Existence for the Pontryagin FBSDE). *Under the combined assumptions of Theorems 2.7 and 2.12, part (1), the Pontryagin FBSDE admits a solution, as soon as the value of the monotone-follower problem is finite.*

Remark 2.15. We do not discuss uniqueness of solutions in detail either in the context of Theorem 2.7 above, or in the context of our other results below. In particular cases, clearly, the strong solution will be unique if enough strict convexity is assumed on the problem ingredients.

2.5. A connection with optimal stopping. In our next result, we revisit, and, more importantly, reinterpret, the celebrated connection between optimal stopping and stochastic control in the context of the generalized monotone-follower problem in dimension $k = 1$. Our formulation of the optimal-stopping problem differs slightly from the classical one, but is easily seen to be essentially equivalent to it (we comment more about it below). It is chosen so as to make our point - namely that the stopping problem associated to the monotone-follower problem is but a manifestation of the maximum principle of Pontryagin - more prominent. It also follows our distributional philosophy and we get to reuse the framework (and the notion) of admissible controls \mathcal{A} from Definition 2.1.

Specifically, we work on the path space $\mathcal{D}_{\cdot, \uparrow}^{d+1}(L, A)$ and, assuming that the functions g and h are continuously-differentiable in a , with derivatives denoted by g_a and h_a , we define

$$(2.3) \quad K(\mathbb{P}) = \mathbb{E}^{\mathbb{P}} \left[\left(f(\tau_A) + g_a(L_{\tau_A}, 0) + \int_{\tau_A}^T h_a(L_t, 0) dt \right) \mathbf{1}_{\{\tau_A < \infty\}} \right] \in \mathbb{R},$$

where τ_A is the stopping time given by

$$\tau_A = \inf\{t \geq 0 : A_t > 0\}, \text{ with } \inf \emptyset = +\infty,$$

whenever the expression inside the expectation in (2.3) above is in $\mathbb{L}^1(\mathbb{P})$; the set of all such \mathbb{P} is denoted by \mathcal{A}_S .

Definition 2.16. A probability $\hat{\mathbb{P}} \in \mathcal{A}_S$ is said to be a **solution** of the optimal-stopping problem if $K(\hat{\mathbb{P}}) \leq K(\mathbb{P})$ for all $\mathbb{P} \in \mathcal{A}_S$.

Remark 2.17. Viewed in isolation, the above formulation of the optimal stopping problem contains obvious redundancies (the \mathbb{P} -behavior of A after τ_A , for example). Even when the class of the probability measures $\mathbb{P} \in \mathcal{A}$ is further restricted so that A becomes a single-jump 0-to-1 process, \mathbb{P} -a.s., our formulation corresponds to a randomized optimal stopping problem, in that A is allowed to depend on innovations independent of L . All in all, part (2) of Definition 2.1 makes the problem equivalent to a randomized optimal stopping problem with respect to the right-continuous augmentation of $\{\mathcal{F}_t^Y\}_{t \in [0, T]}$. There is no harm, however, since it turns out that, as usual in optimal stopping, randomization leads to no increase in value.

Theorem 2.18 (A connection between control and optimal stopping). *Suppose that $k = 1$ and that the assumptions of Theorem 2.12, part (1), hold. Then any solution to the monotone-follower problem is also a solution to the optimal-stopping problem.*

Remark 2.19. As we do not use the notion of a value function, there is no analogue of the equation (3.17) in Theorem 3.4, p. 862 in [KS84] about equality between the derivative (gradient) of the value function in the control problem and the value of the optimal stopping problem. The statements about the relationship between the optimal control in the former and the optimal stopping time in the later translate directly into our setting. The reader will see that the (short) proof of Theorem 2.18

below, given in subsection 3.3, it is nothing but a simple observation, once the Pontryagin principle is established.

2.6. The approximation result. In order to understand the monotone-follower problem better and to provide an approach to it with computation in mind, we pose a sequence of its “capped” versions. These play the role of natural regular approximands to the inherently singular monotone-follower problem. The setting follows closely that of the previous section. The only difference is that the set of allowed controls consists only of Lipschitz-continuous nondecreasing processes, without the initial jump. More precisely, we have the following definition:

Definition 2.20 (Admissible capped controls). Given $n \in \mathbb{N}$, a probability $\mathbb{P} \in \mathfrak{P}_{\cdot, \uparrow}^{d+k}(L, A)$ is called **n -capped admissible**, denoted by $\mathbb{P} \in \mathcal{A}^{[n]}$, if $\mathbb{P} \in \mathcal{A}$ and, \mathbb{P} -a.s., the coordinate process A is Lipschitz continuous with the Lipschitz constant at most n , and $A_0 = 0$, \mathbb{P} -a.s. The **value** of the n -th capped problem is given by

$$V^{[n]} = \inf_{\mathbb{P} \in \mathcal{A}^{[n]}} J(\mathbb{P}),$$

and we say that the probability measure $\hat{\mathbb{P}} \in \mathcal{A}^{[n]}$ is the **weak solution** to the capped monotone-follower problem if $V^{[n]} = J(\hat{\mathbb{P}}) < \infty$.

While Theorem 2.7 relied on a minimal set of assumptions, the approximation result we give below requires more structure. Here, $C_c^\infty(\mathbb{R}^d)$ denote the set of all infinitely-differentiable functions on \mathbb{R}^d with compact support, while $C_b(\mathbb{R}^d)$ refers to the set of all bounded continuous functions; λ denotes the Lebesgue measure on $[0, T]$.

Theorem 2.21 (Approximation by regular controls). *Suppose that*

- (1) *The law \mathbb{P}_0 is Feller, in that for each $t \in [0, T)$*
 - (a) *the σ -algebras \mathcal{F}_{t+}^L and \mathcal{F}_t^L on \mathcal{D}^d coincide \mathbb{P}_0 -a.s.*
 - (b) *for each $G \in C_c^\infty(\mathbb{R}^d)$, there exists $G^* \in C_b(\mathbb{R}^d)$ such that*

$$\mathbb{E}^{\mathbb{P}_0}[G(L_T) | \mathcal{F}_{t+}^L] = G^*(L_t), \quad \mathbb{P}_0\text{-a.s.}$$

- (2) *The coordinate process L is a quasimartingale under \mathbb{P}_0*
- (3) *The primitives f, g and h are regular enough, in that*
 - (a) *each component of f is uniformly bounded away from 0,*
 - (b) *the functions $g(\cdot, a)$ and $h(\cdot, a)$ are continuous for each $a \in [0, \infty)^k$.*
 - (c) *$h(l, \cdot)$, and $g(l, \cdot)$ are continuously differentiable and convex on $[0, \infty)^k$ for each $l \in \mathbb{R}^d$, and there exist $p, q > 1$ and Borel functions $\Phi_g, \Phi_h : \mathbb{R}^d \rightarrow [0, \infty)$ with*

$$\Phi_h(L) \in \mathbb{L}^p(\lambda \otimes \mathbb{P}_0) \text{ and } \Phi_g(L_T) \in \mathbb{L}^p(\mathbb{P}_0),$$

such that, for $\varphi \in \{g, h\}$, we have

$$\varphi(l, 0) + |\nabla \varphi(l, a)| \leq \Phi_\varphi(l) + |a|^q, \text{ for all } (l, a) \in \mathbb{R}^{d+k}.$$

Then

- For each n , the capped problem admits a solution $\hat{\mathbb{P}}^{(n)} \in \mathcal{A}^{[n]}$ and

$$V^{[n]} \searrow V.$$

- A subsequence of the sequence $\{\hat{\mathbb{P}}^{(n)}\}_{n \in \mathbb{N}}$ converges in the Meyer-Zheng sense to a solution $\hat{\mathbb{P}}$ of the monotone follower problem.

Remark 2.22.

- (1) There are several slightly-different classes of processes found under the name of a Feller process in the literature, so we make the essential properties needed in the proof explicit in the statement. These particular properties are, furthermore, implied by all the definitions of the Feller property the authors have encountered. Consequently, all standard examples of Feller processes such as diffusions, stable processes, Lévy processes, etc., fall under our framework.
- (2) The quasimartingality assumption on L is put in place mostly for convenience. It is known that so-called “nice” Feller processes (the domain of whose generator contains smooth functions with compact support) are automatically special semimartingales and, therefore, local quasimartingales (see [Sch12] for the first part of the statement, and [Kal02, Theorem 23.20, p. 451] for the second). As no convexity in the variable l is assumed, one can further do away with the localization in many cases by replacing L by $q(L)$, where q is a smooth, injective and bounded function. Such a replacement would not change the problem; indeed, conditions (1) and (3) of Theorem 2.21 are invariant under the transformation $L \mapsto q(L)$.
- (3) The growth assumptions on the functions f, g and h are essentially those of [KS84], rephrased in our language. We note the fact that f is bounded away from zero immediately implies the linear coercivity condition of Theorem 2.7, while the condition $\varphi(l, 0) \leq \Phi_\varphi(l)$, for $\varphi \in \{g, h\}$, guarantees that the value is finite.

Example 2.23. In general, the sequence of capped optimizers cannot be guaranteed to converge towards a minimizer \mathbb{P}^* weakly, under the Skorokhod topology. Indeed, Skorokhod convergence preserves continuity, and all capped optimal controls are continuous, but it is easily seen that the solution to the monotone-follower problem does not need to be a continuous process. Indeed, it suffices to take $k = d = 1$, any \mathbb{P}_0 with $\mathbb{P}_0[L_T > 1] > 0$, $f \equiv 1$, $h \equiv 0$ and $g(l, a) = \frac{1}{2}(l - a)^2$, so that the optimal A is given by $A_t = 0$ for $t < T$ and $A_T = \max(0, L_T - 1)$.

On the other hand, if one can guarantee that the optimizer is continuous (and $A_0 = 0$), the Meyer-Zheng convergence automatically upgrades to the weak convergence in $C[0, T]$ (see [Pra99]).

One of the immediate consequences of Theorem 2.21 is that the monotone-follower problem can be posed over Lipschitz controls, without affecting the value function.

Corollary 2.24 (Lipschitz ε -optimal controls). *Under the conditions of Theorem 2.21 for each $\varepsilon > 0$ there exists $M > 0$ and an ε -optimal admissible control \mathbb{P} , such that A is uniformly M -Lipschitz, \mathbb{P} -a.s.*

3. PROOFS

Proofs of our main results, namely Theorems 2.7, 2.12, 2.18 and 2.21 are collected in this section. The proof of each theorem occupies a section of its own, and all the conditions stated in the theorem are assumed to hold - without explicit mention - throughout the section.

3.1. A proof of Theorem 2.7. We start with an auxiliary result which states that an admissible control can always be turned into a strong admissible control without any sacrifice in value. The central idea is that, even though the optional projection of a nondecreasing process is *not necessarily nondecreasing* in general, this turns out to be so in our setting.

Proposition 3.1. *For $\mathbb{P} \in \mathcal{A}$ with $\mathbb{E}^\mathbb{P}[A_T] < \infty$ let ${}^\circ A$ be the optional projection of A onto the right-continuous and complete augmentation $\bar{\mathbb{F}}_+^L$ of the natural filtration \mathbb{F}^L . Then the joint law ${}^\circ \mathbb{P}$ of $(L, {}^\circ A)$ is admissible and $J({}^\circ \mathbb{P}) \leq J(\mathbb{P})$.*

Proof. The optional projection of a càdlàg process onto a filtration satisfying the usual conditions is indistinguishable from a càdlàg process (see, e.g., Theorem 2.9, p. 18 in [BC09]). It is an immediate consequence of the condition (2) of Definition 2.1 that

$$\mathbb{E}^\mathbb{P}[A_t | \bar{\mathcal{F}}_{t+}^L] = \mathbb{E}^\mathbb{P}[A_t | \mathcal{F}_T^L], \text{ a.s., for all } t \in [0, T],$$

and, so ${}^\circ A_t = \mathbb{E}^\mathbb{P}[A_t | \mathcal{F}_T^L] \leq \mathbb{E}^\mathbb{P}[A_s | \mathcal{F}_T^L] = {}^\circ A_s$, a.s., for $s \leq t$. By construction, the σ -algebras $\bar{\mathcal{F}}_{t+}^L$ and \mathcal{F}_{t+}^L differ only in \mathbb{P} -negligible sets, and, so, $\bar{\mathcal{F}}_T$ and $\bar{\mathcal{F}}_{t+}^L$ are conditionally independent given \mathcal{F}_{t+}^L , which, in turn, implies that the joint law of (L, A) is admissible.

Next, we show that $J({}^\circ A) \leq J(A)$. For $\varphi \in \{g, h\}$ we denote by $\tilde{\varphi}(l, \cdot)$ the convex conjugate (in the second variable) of φ :

$$\tilde{\varphi}(l, \alpha) = \sup_{a \geq 0} (\alpha a - \varphi(l, a)) \text{ so that } \varphi(l, a) = \sup_{\alpha \in \mathbb{R}} (\alpha a - \tilde{\varphi}(l, \alpha)).$$

Then, for any bounded \mathcal{F}_T^L -measurable random variable α_T with $\tilde{\varphi}(L_t, \alpha_T) < \infty$, \mathbb{P} -a.s., we have

$$\mathbb{E}^\mathbb{P}[\varphi(L_t, A_t) | \mathcal{F}_T^L] \geq \mathbb{E}^\mathbb{P}[\alpha_T A_t | \mathcal{F}_T^L] - \tilde{\varphi}(L_t, \alpha_T) = \alpha_T {}^\circ A_t - \tilde{\varphi}(L_t, \alpha_T), \text{ } \mathbb{P}\text{-a.s.}$$

The \mathbb{P} -essential supremum of the right-hand side over all bounded \mathcal{F}_T^L -measurable α_T is easily seen to be equal to $\varphi(L_t, {}^\circ A_t)$, \mathbb{P} -a.s., for $t \in [0, T]$, so, by the tower property, $\mathbb{E}^\mathbb{P}[\varphi(L_t, {}^\circ A_t)] \leq \mathbb{E}^\mathbb{P}[\varphi(L_t, A_t)]$. Thus,

$$\mathbb{E}^\mathbb{P}\left[\int_0^T h(L_t, {}^\circ A_t) dt + g(L_T, {}^\circ A_T)\right] \leq \mathbb{E}^\mathbb{P}\left[\int_0^T h(L_t, A_t) dt + g(L_T, A_T)\right].$$

Finally, we let \mathcal{M} denote the set of all bounded measurable functions $\psi : [0, T] \rightarrow \mathbb{R}$ with

$$(3.1) \quad \mathbb{E}\left[\int_{[0, T]} \psi(t) d{}^\circ A_t\right] = \mathbb{E}\left[\int_{[0, T]} \psi(t) dA_t\right].$$

\mathcal{M} is clearly a monotone class which contains all functions of the form $\psi(t) = \mathbf{1}_{(a, T]}(t)$, so, by the monotone-class theorem, it contains all bounded measurable functions and, in particular, f . \square

Continuing with the proof of Theorem 2.7, we assume that its value is finite, pick a minimizing sequence $\{\mathbb{P}_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$, and use it to build a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and, on it, the sequence $L, A^{(1)}, A^{(2)}, \dots$, as in Lemma A.2.

Thanks to Proposition 3.1, we may assume, without loss of generality, that all $A^{(n)}$ are $\bar{\mathbb{F}}_+^L$ -adapted, where $\bar{\mathbb{F}}_+^L = \{\bar{\mathcal{F}}_{t+}^L\}_{t \in [0, T]}$ denotes the right-continuous and complete augmentation of the natural filtration $\{\mathcal{F}_t^L\}_{t \in [0, T]}$,

Now that a common probability space has been constructed, we follow the methodology of [BR01] and [RS11]. Thanks to the linear coercivity condition (2.1), the sequence $\{A_T^{(n)}\}_{n \in \mathbb{N}}$ is bounded in \mathbb{L}^1 ; also, all $A^{(n)}$ are $\bar{\mathbb{F}}_+^L$ -adapted, and $\bar{\mathbb{F}}_+^L$ is right-continuous. Therefore, we can use Lemma 3.5, p. 470, in [Kab99] to guarantee the existence of an $\bar{\mathbb{F}}_+^L$ -adapted process B , with paths in \mathcal{D}^d and a sequence $\{B^{(n)}\}_{n \in \mathbb{N}}$ of Cesàro means of a subsequence of $\{A_n^{(n)}\}_{n \in \mathbb{N}}$ which converges to B in the following sense (the sense of optional random measures): for almost all ω , the Stieltjes measures induced by $B^{(n)}(\omega)$ converge weakly towards to the Stieltjes measure induced by $B(\omega)$. In particular, there exists a countable subset \mathcal{N} of $[0, T)$ (the set of jumps of $t \mapsto \mathbb{E}[B_t]$ on $[0, T)$) such that

$$\int_0^T f(t) dB_t^{(n)} \rightarrow \int_0^T f(t) dB_t, \text{ a.s., and } B_t^{(n)} \rightarrow B_t, \text{ a.s., for all } t \in [0, T] \setminus \mathcal{N}.$$

Therefore, by Fatou's lemma (applied on Ω for the first and the third term, and on the product space $[0, T] \times \Omega$ for the second), we have

$$\begin{aligned} \mathbb{E}\left[\int_0^T f(t) dB_t + \int_0^T h(L_t, B_t) dt + g(L_T, B_T)\right] &\leq \\ &\leq \liminf_{n \rightarrow \infty} \mathbb{E}\left[\int_0^T f(t) dB_t^{(n)} + \int_0^T h(L_t, B_t^{(n)}) dt + g(L_T, B_T^{(n)})\right]. \end{aligned}$$

For a nondecreasing càdlàg process A on $(\Omega, \mathcal{F}, \mathbb{P})$ we set $J(A) = \mathbb{E}[C(L, A)]$ and notice that the convexity of J and the fact that $J(A^{(n)}) \searrow \inf_{\mathbb{P} \in \mathcal{A}} J(\mathbb{P})$ together yield that $\{B^{(n)}\}_{n \in \mathbb{N}}$ is a minimizing sequence, too, in that $J(B^{(n)}) \searrow \inf_{\mathbb{P} \in \mathcal{A}} J(\mathbb{P})$. Therefore, $J(B) \leq \inf_{\mathbb{P} \in \mathcal{A}} J(\mathbb{P})$ and it only remains to note that the law of (L, B) is strongly admissible since B is $\bar{\mathbb{F}}_+^L$ -adapted.

3.2. A proof of Theorem 2.12. To streamline the presentation in this and the subsequent subsections, we introduce additional notation: the **subgradient** map $\partial C(L, A) : [0, T] \rightarrow \mathbb{R}^k$, at $(L, A) \in \mathcal{D}_{\cdot, \uparrow}^{d+k}$, is given by

$$\partial C(L, A)_t = f(t) + \int_t^T \nabla h(L_s, A_s) ds + \nabla g(L_T, A_T) \text{ for } t \in [0, T],$$

where, as usual, ∇h and ∇g denote the gradients with respect to the second variable. The reader will easily check that $\partial C(L, A)$ has the following property (which earns it the name subgradient):

$$(3.2) \quad C(L, A + \Delta) \geq C(L, A) + \langle \partial C(L, A), \Delta \rangle,$$

for all $\Delta \in \mathcal{D}_{fv}^k$ with $A + \Delta \in \mathcal{D}_{\uparrow}^k$, where

$$\langle X, \Delta \rangle = \int_{[0, T]} X_u d\Delta_u.$$

We also note, for future reference and using integration by parts, that

$$(3.3) \quad \langle \partial C(L, A), \Delta \rangle = \int_{[0, T]} f(t) d\Delta_t + \int_0^T \nabla h(L_t, A_t) \Delta_t dt + \nabla g(L_T, A_T) \Delta_T,$$

for all $\Delta \in \mathcal{D}_{fv}^k$.

We start the proof by assuming that $\hat{\mathbb{P}} \in \mathfrak{P}^{d+k}(L, A)$ solves the monotone-follower problem, with value $V = J(\hat{\mathbb{P}}) < \infty$. In particular, we have $C(L, A) \in \mathbb{L}^1(\hat{\mathbb{P}})$. To relieve the notation we work on the sample space $\Omega = \mathcal{D}_{\uparrow}^{d+k}(L, A)$, under the probability $\hat{\mathbb{P}}$, until the end of this part of the proof. Moreover, thanks to assumptions of the theorem, for $\varphi \in \{g, h\}$, $l \in \mathbb{R}^k$, $a \in [0, \infty)^d$ and $x \in \mathbb{R}^d$ such that $a + x \in [0, \infty)^d$, we have, for each $c \in (0, 1)$,

$$\begin{aligned} |\nabla \varphi(l, a + cx)| &\leq \Phi_\varphi(l) + M\varphi(l, a + cx) \\ &= \Phi_\varphi(l) + M\varphi(l, a) + M \int_0^c \langle x, \nabla \varphi(l, a + tx) \rangle dt \\ &\leq \left(\Phi_\varphi(l) + M\varphi(l, a) \right) + M|x| \int_0^c |\nabla \varphi(l, a + tx)| dt \end{aligned}$$

Gronwall's inequality then implies that

$$(3.4) \quad |\nabla \varphi(l, a + x)| \leq \left(\Phi_\varphi(l) + \varphi(l, a) \right) e^{M|x|}.$$

Let \mathcal{V}_A denote the set of all bounded processes Δ with paths in \mathcal{D}_{fv}^k , adapted to the natural filtration $\mathbb{F}^{L, A}$ such that,

$$\text{either } \Delta \in \mathcal{D}_{\uparrow}^k \text{ or } \Delta = -\frac{1}{2} \min(A, n) \text{ for some } n \in \mathbb{N}.$$

It has the property that for $\varepsilon \in [0, 1]$ and $\Delta \in \mathcal{V}_A$, the joint law \mathbb{P}^ε of (L, A^ε) , where $A^\varepsilon = A + \varepsilon \Delta$, is an admissible probability measure in \mathfrak{P}^{d+k} . By the optimality of A and (3.2), we have

$$\mathbb{E}[C(L, A)] \leq \mathbb{E}[C(L, A^\varepsilon)] \leq \mathbb{E}[C(L, A) + \langle \partial C(L, A^\varepsilon), \varepsilon \Delta \rangle],$$

from where it follows that

$$(3.5) \quad \langle \partial C(L, A^\varepsilon), \Delta \rangle^- \in \mathbb{L}^1 \text{ and } \mathbb{E}[\langle \partial C(L, A^\varepsilon), \Delta \rangle] \geq 0, \text{ for all } \varepsilon \in [0, 1].$$

Thanks to boundedness of processes in \mathcal{V}_A and the fact that $C(L, A)$ is integrable, the inequality (3.4) implies that the family

$$\left\{ \langle \partial C(L_t, A_t^\varepsilon), \Delta \rangle : \varepsilon \in [0, 1] \right\} \text{ is uniformly integrable for all } \Delta \in \mathcal{V}_A.$$

Moreover, both ∇h and ∇g are continuous, so

$$\lim_{\varepsilon \rightarrow 0} \langle \partial C(L, A^\varepsilon), \Delta \rangle = \langle \partial C(L, A), \Delta \rangle, \text{ a.s.}$$

It follows that we can pass to the limit as $\varepsilon \rightarrow 0$ in (3.5) to conclude that

$$(3.6) \quad \mathbb{E}[\langle \partial C(L, A), \Delta \rangle] \geq 0, \text{ for all } \Delta \in \mathcal{V}_A,$$

and, consequently, that

$$(3.7) \quad \mathbb{E}[\langle Y, \Delta \rangle] \geq 0, \text{ for all } \Delta \in \mathcal{V}_A,$$

where Y denotes the optional projection of $\partial C(L, A)$ onto the right-continuous and complete augmentation $\bar{\mathbb{F}}_+^{L, A}$ of $\mathbb{F}^{L, A}$. Since $\partial C(L, A)$ is càdlàg, the process Y can be chosen in a càdlàg version,

too (see Theorem 2.9, p. 18 in [BC09]). Hence, by varying Δ in the class of nondecreasing processes in \mathcal{V}_A , we can conclude that $Y_t \geq 0$, for all $t \in [0, T]$, a.s.

On the other hand if we use each element of the sequence $\Delta_n = -\frac{1}{2} \min(A, n)$ in (3.7), we obtain

$$\int_{[0, T]} Y_t dA_t = 0, \text{ a.s..}$$

In order to show that the law $\tilde{\mathbb{P}}$ of the triple (L, A, Y) solves the Pontryagin FBSDE, we only need to argue that $Y + \int_0^\cdot \nabla h(L_t, A_t) dt - f$ is an $\mathbb{F}^{L, A, Y}$ martingale (under $\tilde{\mathbb{P}}$, on \mathcal{D}^{d+2k}). This follows directly from the fact that Y is a càdlàg version of the optional projection of $\partial C(L, A)$ onto $\mathbb{F}^{L, A, Y}$.

Conversely, let $\tilde{\mathbb{P}} \in \mathfrak{P}^{d+2k}(L, A, Y)$ be a solution to the Pontryagin FBSDE. To prove that $\hat{\mathbb{P}} = \tilde{\mathbb{P}}_{L, A}$ is a weak minimizer in the monotone-follower problem, we pick a competing admissible measure $\mathbb{P}' \in \mathcal{A}$. Using Lemma A.1, we construct the measure $\mathbb{P} = \tilde{\mathbb{P}} \otimes \mathbb{P}'$ on \mathcal{D}^{d+3k} (with coordinates (L, A, Y, A')). Since $\mathbb{P}_{(L, A, Y)}$ solves the Pontryagin FBSDE, $Y + \int_0^\cdot \nabla h(L_t, A_t) dt - f$ is an $(\mathbb{F}^{L, A, Y}, \mathbb{P})$ -martingale. Moreover, the L -conditional independence between A' and (A, Y) implies that it is also an $(\mathbb{F}^{L, A, Y, A'}, \mathbb{P})$ -martingale. Consequently, we have

$$\mathbb{E}^\mathbb{P} [\langle \partial C(L, A), A' \rangle] = \mathbb{E}^\mathbb{P} [\langle Y, A' \rangle] \text{ and } \mathbb{E}^\mathbb{P} [\langle \partial C(L, A), A \rangle] = \mathbb{E}^\mathbb{P} [\langle Y, A \rangle].$$

The subgradient identity (3.2) then implies that

$$\begin{aligned} (3.8) \quad J(\mathbb{P}') &= \mathbb{E}^\mathbb{P} [C(L, A')] \geq \mathbb{E}^\mathbb{P} [C(L, A) + \langle \partial C(L, A), A' - A \rangle] \\ &= J(\hat{\mathbb{P}}) + \mathbb{E}^\mathbb{P} [\langle Y, A' - A \rangle] = J(\hat{\mathbb{P}}) + \mathbb{E}^\mathbb{P} [\langle Y, A' \rangle] \geq J(\hat{\mathbb{P}}). \end{aligned}$$

3.3. A proof of Theorem 2.18. Let $\hat{\mathbb{P}} \in \mathcal{A}$ be a solution to the monotone-follower problem. By Theorem 2.12, part (1), it can be realized as the marginal $\tilde{\mathbb{P}}_{L, A}$ of some solution $\tilde{\mathbb{P}}_{L, A, Y}$ of the Pontryagin FBSDE. For an admissible measure $\mathbb{P}' \in \mathcal{A}$, and using Lemma A.1, we can construct the measure $\mathbb{P} = \tilde{\mathbb{P}} \otimes \mathbb{P}'$ on $\mathcal{D}_{\uparrow, \uparrow}^{d+3}$ (with coordinates (L, A, Y, A')) and work on $\mathcal{D}_{\uparrow, \uparrow}^{d+3}$ under \mathbb{P} for the remainder of the proof. As argued in the previous subsection, the process $Y + \int_0^\cdot h_a(L_t, A_t) dt - f$ is an $(\mathbb{F}^{L, A, Y, A'}, \mathbb{P})$ -martingale, and, so,

$$\mathbb{E}[Y_{\tau_{A'}} \mathbf{1}_{\{\tau_{A'} < \infty\}}] = \mathbb{E}[\partial C(L, A)_{\tau_{A'}} \mathbf{1}_{\{\tau_{A'} < \infty\}}]$$

where $\tau_{A'} = \inf\{t \geq 0 : A'_t > 0\} \in [0, T] \cup \{\infty\}$. By the assumptions of convexity we placed on h and g , we have the following inequalities

$$h_a(L_s, 0) - h_a(L_s, A_s) \leq 0 \text{ and } g_a(L_T, 0) - g_a(L_T, A_T) \leq 0,$$

for all $s \in [0, T]$, a.s. Therefore, by the nonnegativity of Y , we have

$$\begin{aligned} K(\mathbb{P}') &= \mathbb{E}[\partial C(L, 0)_{\tau_{A'}} \mathbf{1}_{\{\tau_{A'} < \infty\}}] \geq \mathbb{E}[\partial C(L, 0)_{\tau_{A'}} \mathbf{1}_{\{\tau_{A'} < \infty\}} - Y_{\tau_{A'}} \mathbf{1}_{\{\tau_{A'} < \infty\}}] \\ &= \mathbb{E}\left[\int_{\tau_{A'}}^T (h_a(L_s, 0) - h_a(L_s, A_s)) ds + (g_a(L_T, 0) - g_a(L_T, A_T)) \mathbf{1}_{\{\tau_{A'} < \infty\}}\right] \\ &\geq \mathbb{E}\left[\int_0^T (h_a(L_s, 0) - h_a(L_s, A_s)) ds + (g_a(L_T, 0) - g_a(L_T, A_T))\right] \\ &= \mathbb{E}[\partial C(L, 0)_0 - Y_0] \end{aligned}$$

On the other hand, if we repeat the computation above with $\tau_{A'}$ replaced by τ_A , all the inequalities become equalities, implying that $K(\mathbb{P}) \leq K(\mathbb{P}')$. Indeed, we clearly have

$$h_a(L_s, 0) = h_a(L_s, A_s), \text{ on } \{s < \tau_A\},$$

and

$$g_a(L_T, 0) = g_a(L_T, A_T) \text{ on } \{\tau_A = \infty\},$$

as well as

$$\mathbb{E}[Y_{\tau_A} \mathbf{1}_{\{\tau_A < \infty\}}] = 0,$$

where this last equality follows from the fact that $\int_0^T Y_u dA_u = 0$.

3.4. A proof of Theorem 2.21. We start by posing the capped monotone-follower problems on a common fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which hosts a càdlàg process L with distribution \mathbb{P}_L , and consider only right-continuous and complete augmentation $\bar{\mathbb{F}}_+^L$ of the natural filtration \mathbb{F}^L , generated by L . Let $\mathcal{U}^{[n]}$ denote the set of all progressively-measurable k -dimensional processes with values in $[0, n]^k$. For $u \in \mathcal{U}^{[n]}$, all components of the process $A = \int_0^\cdot u(t) dt$ are Lipschitz continuous with the Lipschitz constant not exceeding n . Conversely, each adapted process with such Lipschitz paths admits a similar representation. This correspondence allows us to pose the n -th capped monotone follower problem either over the set of process $\mathcal{U}^{[n]}$ or over the appropriate admissible set $\mathcal{A}^{[n]} = \{\int_0^\cdot u(t) dt : u \in \mathcal{U}^{[n]}\}$. Their (strong) value functions are then defined by

$$(3.9) \quad \tilde{V}^{[n]} = \inf_{A \in \mathcal{A}^{[n]}} \mathbb{E}[C(L, A)] = \inf_{u \in \mathcal{U}^{[n]}} J(u) \text{ where } J(u) = \mathbb{E}[C(L, \int_0^\cdot u)].$$

Each $A \in \mathcal{A}^{[n]}$ is $\bar{\mathcal{F}}_+^L$ -adapted and, therefore, strongly admissible, in the sense of Definition 2.20. In particular, $\tilde{V}^{[n]} \geq V$, for all n . Also, noting that the polynomial-growth assumption implies that $\mathbb{E}[C(L, A)] < \infty$, for each bounded A , we have $\tilde{V}^{[n]} < \infty$, for all $n \in \mathbb{N}$, and, consequently, $V < \infty$.

For readability, we split the remainder of the proof into several subsections.

3.4.1. Existence in the prelimit. Let $\mathbb{L}^2([0, T] \times \Omega, \text{Prog})$ denote the space of all $(\lambda \otimes \mathbb{P})$ -equivalence classes) of $\bar{\mathbb{F}}_+^L$ -progressively-measurable processes u on $[0, T] \times \mathbb{P}$ with

$$\|u\|_{\mathbb{L}^2([0, T] \times \Omega, \text{Prog})} = \mathbb{E} \left[\int_0^T |u(t)|^2 dt \right]^{1/2} < \infty.$$

Proposition 3.2. *The infimum in (3.9) is attained at some $u^{[n]} \in \mathcal{U}^{[n]}$.*

Proof. We proceed in the standard way, using the so-called “direct method”. Let $\{u_k\}_{k \in \mathbb{N}} \subset \mathcal{U}^{[n]}$ be a minimizing sequence, i.e., $J(u_k) \searrow \tilde{V}^{[n]}$. Since $\mathcal{U}^{[n]}$ is bounded in $\mathbb{L}^2([0, T] \times \Omega, \text{Prog})$, the Banach-Sachs theorem implies that we can extract a subsequence whose Cesàro sums (still denoted by $\{u_k\}_{k \in \mathbb{N}}$) converge strongly towards some $u^{[n]} \in \mathbb{L}^2([0, T] \times \Omega, \text{Prog})$. Furthermore, given that $\mathcal{U}^{[n]}$ is closed and convex, we have $u^{[n]} \in \mathcal{U}^{[n]}$, as well. Thanks to the convexity of J , which is inherited from C , $\{u_k\}_{k \in \mathbb{N}}$ remains a minimizing sequence. Hence, to show that $u^{[n]}$ is the minimizer, it will be enough to establish lower semicontinuity of J on $\mathcal{U}^{[n]}$ which is, in turn, a direct consequence of Fatou’s lemma. \square

3.4.2. *A version of the Pontryagin FBSDE.* Having established the existence in the (strong) capped monotone follower problem, for each $n \in \mathbb{N}$ we pick and fix a minimizer $u^{[n]}$ as in Proposition 3.2 and turn to a capped version of the Pontryagin FBSDE. We state it in a very weak form (namely, as Proposition 3.3) which will, nevertheless suffice to establish the validity of the full Pontryagin FBSDE in the limit. The following notation will be used throughout:

$$A_t^{[n]} = \int_0^t u_s^{[n]} ds, \quad N_t^{[n]} = \int_0^t \nabla h(L_s, A_s^{[n]}) ds, \quad F_t^{[n]} = \int_0^t f(s) dA_s^{[n]},$$

as well as

$$M_t^{[n]} = \mathbb{E} \left[\nabla g(L_T, A_T^{[n]}) + N_T^{[n]} \middle| \bar{\mathcal{F}}_{t+}^L \right], \quad Y_t^{[n]} = f(t) + M_t^{[n]} - N_t^{[n]},$$

all taken in their càdlàg versions. We note immediately that, thanks to the polynomial-growth condition, all the integrals above are well defined, and that $Y^{[n]}$ is the optional projection of $\partial C_A(L, A^{[n]})$ onto $\bar{\mathbb{F}}_+^L$.

Proposition 3.3. *For $n \in \mathbb{N}$, we have*

$$(3.10) \quad n \mathbb{E} \left[\int_0^T (Y_t^{[n]})^- dt \right] = - \mathbb{E} \left[\int_0^T Y_t^{[n]} dA_t^{[n]} \right]$$

and

$$(3.11) \quad \lim_{n \rightarrow \infty} \mathbb{E} \left[\int_0^T (Y_t^{[n]})^- dt \right] = 0.$$

Proof. Given $v \in \mathcal{U}^{[n]}$ and $\varepsilon \in [0, 1]$ we set $B = \int_0^\cdot v_t dt$ and define

$$A^\varepsilon = A^{[n]} + \varepsilon(B - A^{[n]}) \in \mathcal{A}^{[n]}$$

Since $C(L, A^{[n]}) \in \mathbb{L}^1$, the optimality of $u^{[n]}$ implies that

$$0 \geq \mathbb{E}[C(L, A^{[n]})] - \mathbb{E}[C(L, A^\varepsilon)] \geq \varepsilon \mathbb{E}[\langle \partial C(L, A^\varepsilon), A^{[n]} - B \rangle],$$

We let $\varepsilon \searrow 0$ and use the dominated convergence theorem to conclude that

$$(3.12) \quad \mathbb{E} \left[\int_0^T (Y_t^{[n]})^+ (u_t^{[n]} - v_t) dt \right] \leq \mathbb{E} \left[\int_0^T (Y_t^{[n]})^- (u_t^{[n]} - v_t) dt \right].$$

Setting $v = n \mathbf{1}_{\{Y^{[n]} \leq 0\}}$ yields

$$(3.13) \quad \mathbb{E} \left[\int_0^T (Y_t^{[n]})^+ u_t^{[n]} dt \right] \leq \mathbb{E} \left[\int_0^T (Y_t^{[n]})^- (u_t^{[n]} - n) dt \right].$$

Since the left-hand side of (3.13) is nonnegative and the right-hand side nonpositive, we conclude that both of them vanish, which, in turn, directly implies (3.10).

To show (3.11) we use the inherited subgradient property of $Y^{[n]}$ and (3.10) to obtain

$$\begin{aligned} 0 &\leq \mathbb{E} [C(L, A^{[n]})] \leq \mathbb{E} [C(L, 0)] + \mathbb{E} \left[\int_0^T Y_u^{[n]} dA_u^{[n]} \right] \\ &= \mathbb{E} [C(L, 0)] - n \mathbb{E} \left[\int_0^T (Y_t^{[n]})^- dt \right]. \end{aligned}$$

□

3.4.3. *Relative compactness in the Meyer-Zheng topology.* Our next step is to pass to the limit, as $n \rightarrow \infty$, in the Meyer-Zheng convergence and show that the limiting law satisfies the weak FBSDE (2.10). The reader will find a short recapitulation of the pertinent known results on the Meyer-Zheng convergence (minimally modified to fit our needs) in subsections A.3, A.4 and A.5 of Appendix A.

In the sequel, $\{\tilde{\mathbb{P}}^{(n)}\}_{n \in \mathbb{N}}$ denotes the sequence of laws of the triplets $(L, A^{[n]}, M^{[n]})$ on \mathcal{D}^{d+2k} .

Proposition 3.4. *For each $p \geq 1$, we have*

$$(3.14) \quad \sup_n \|A_T^{[n]}\|_{\mathbb{L}^p} < \infty,$$

and the sequence $\{\tilde{\mathbb{P}}^{(n)}\}_{n \in \mathbb{N}}$ is relatively compact in the Meyer-Zheng topology on \mathfrak{P}^{d+2k} .

Proof. Since the distribution of first component L does not depend on n , by Theorem A.5, it will be enough to establish that

$$\sup_{n \in \mathbb{N}} \text{Var}^{\mathbb{P}^n}[A] < \infty \text{ and } \sup_{n \in \mathbb{N}} \text{Var}^{\mathbb{P}^n}[M] < \infty,$$

where $\text{Var}^{\mathbb{P}^n}$ denotes the conditional variation (in the quasimartingale sense, as defined in (A.3), below). Moreover, given that all $A^{[n]}$ are nondecreasing, and all $M^{[n]}$ are martingales, relative compactness will follow once we show that

$$\sup_n \mathbb{E}[A_T^{[n]}] < \infty \text{ and } \sup_n \mathbb{E}[|M_T^{[n]}|] < \infty,$$

for which - thanks to our polynomial-growth assumption - it will suffice to establish (3.14). In order to do that, for $n \in \mathbb{N}$ and $r \geq 0$ define $u_t^{[n];r} = u_t^{[n]} 1_{\{A_t^{[n]} < r\}}$, so that

$$A_t^{[n];r} = \int_0^t u_s^{[n];r} ds = A_{t \wedge T^{[n]}(r)}^{[n]},$$

where $T^{[n]}(r) = \inf\{t \in [0, T] : A_t^{[n]} \geq r\} \in [0, T] \cup \{\infty\}$. By the sub-optimality of $u^{[n];r}$ we have

$$\begin{aligned} \mathbb{E} \left[\int_0^T f(t) u_t^{[n];r} dt + \int_0^T h(L_t, A_t^{[n];r}) dt + g(L_t, A_T^{[n];r}) \right] \\ \geq \mathbb{E} \left[\int_0^T f(t) u_t^{[n]} dt + \int_0^T h(L_t, A_t^{[n]}) dt + g(L_t, A_T^{[n]}) \right], \end{aligned}$$

so that

$$\begin{aligned} \mathbb{E} \left[\int_{T \wedge T^{[n]}(r)}^T f(t) u_t^{[n]} dt \right] \\ \leq \mathbb{E} \left[\int_{T \wedge T^{[n]}(r)}^T h(L_t, r) - h(L_t, A_t^{[n]}) dt + \left(g(L_t, r) - g(L_t, A_T^{[n]}) \right) 1_{\{A_T^{[n]} > r\}} \right]. \end{aligned}$$

Since f is positive and componentwise bounded away from zero (say, by $c > 0$), and h, g are nonnegative and convex in their second argument, we have

$$\mathbb{E} \left[\int_{T \wedge T^{[n]}(r)}^T f(t) u_t^{[n]} dt \right] \geq c \mathbb{E} \left[(A_T^{[n]} - r) 1_{\{A_T^{[n]} > r\}} \right],$$

as well as, on $\{A_T^{[n]} > r\}$,

$$\begin{aligned} \int_{T \wedge T^{[n]}(r)}^T h(L_t, r) &\leq \int_{T \wedge T^{[n]}(r)}^T h(L_t, A_t^{[n]}) dt + \int_0^T h(L_t, 0) dt \text{ and} \\ g(L_t, r) &\leq g(L_t, A_T^{[n]}) + g(L_t, 0) \end{aligned}$$

It remains to apply Lemma A.3 with $X = |A_T^{[n]}|$ and $Y = \int_0^T h(L_t, 0) dt + g(L_t, 0)$, to conclude that $\{A_T^{[n]}\}_{n \in \mathbb{N}}$ is bounded in \mathbb{L}^p , for each $p \geq 0$. \square

3.4.4. The Meyer-Zheng limit and its first properties. Having established the relative compactness of the sequence $\{\tilde{\mathbb{P}}^{(n)}\}_{n \in \mathbb{N}}$, we select one of its limit points $\tilde{\mathbb{P}}^*$. By passing to a subsequence, if necessary, we may assume that $\tilde{\mathbb{P}}^{(n)} \rightarrow \tilde{\mathbb{P}}^*$ in the Meyer-Zheng topology.

Proposition 3.5. $\tilde{\mathbb{P}}_{(L,A)}^*$ is (weakly) admissible.

Proof. Since the first components L have the same law under each $\tilde{\mathbb{P}}^{(n)}$ (namely \mathbb{P}_0), it is clear that the same remains true in the limit. To establish the requirement (2) of Definition 2.1, we pick $m \in \mathbb{N}$, two continuous and bounded functions $F : (\mathbb{R}^k)^m \rightarrow \mathbb{R}$ and $H : (\mathbb{R}^d)^m \rightarrow \mathbb{R}$, as well as a $C_c^\infty(\mathbb{R}^d)$ -function G . Thanks to the admissibility of each $\tilde{\mathbb{P}}^{(n)}$, for each $n \in \mathbb{N}$ and all $t < s_1 < \dots < s_m \leq T$, we have

$$\mathbb{E}^{\tilde{\mathbb{P}}^{(n)}} [F G(L_T) | \mathcal{F}_{t+}^L] = \mathbb{E}^{\tilde{\mathbb{P}}^{(n)}} [F | \mathcal{F}_{t+}^L] \mathbb{E}^{\tilde{\mathbb{P}}^{(n)}} [G(L_T) | \mathcal{F}_{t+}^L],$$

where $F = F(A_{s_1}, \dots, A_{s_m})$. Since $\tilde{\mathbb{P}}_L^{(n)} = \mathbb{P}_L$ and thanks to first assumption of Theorem 2.21, for all $n \in \mathbb{N}$ we have

$$\mathbb{E}^{\tilde{\mathbb{P}}^{(n)}} [G(L_T) | \mathcal{F}_{t+}^L] = G^*(L_t), \tilde{\mathbb{P}}^{(n)} - \text{a.s.},$$

for some $G^* \in C_b(\mathbb{R}^k)$. Thus, for $0 \leq r_1 < \dots < r_m \leq t$, we have

$$\mathbb{E}^{\tilde{\mathbb{P}}^{(n)}} [F G(L_T) H] = \mathbb{E}^{\tilde{\mathbb{P}}^{(n)}} [F G^*(L_t) H], \tilde{\mathbb{P}}^{(n)} - \text{a.s.},$$

where $H = H(L_{r_1}, \dots, L_{r_m})$. Thanks to Theorem A.4, after another passage to a subsequence, there exists a full-measure subset \mathcal{T} of $[0, T]$, which includes T , such that \mathbb{P}^n -finite-dimensional distributions with indices in \mathcal{T} converge towards the \mathbb{P} -finite-dimensional distributions. Hence, if $r_1 < \dots < r_m, t$ and $s_1 < \dots < s_m$ belong to such \mathcal{T} , we have

$$\mathbb{E}^{\tilde{\mathbb{P}}^*} [F G(L_T) H] = \mathbb{E}^{\tilde{\mathbb{P}}^*} [F G^*(L_t) H].$$

It follows that, for $t \in \mathcal{T}$, we have

$$(3.15) \quad \mathbb{E}^{\tilde{\mathbb{P}}^*} [F G(L_T) | \mathcal{F}_t^L] = \mathbb{E}^{\tilde{\mathbb{P}}^*} [F | \mathcal{F}_t^L] \mathbb{E}^{\tilde{\mathbb{P}}^*} [G(L_T) | \mathcal{F}_t^L], \tilde{\mathbb{P}}^* - \text{a.s.},$$

for all F, G . It is a part of our assumptions that a version of the Blumenthal's 0 – 1-law holds. By Proposition 3.5, $\tilde{\mathbb{P}}_L^* = \mathbb{P}_L^0$; it follows that σ -algebras \mathcal{F}_t^L and \mathcal{F}_{t+}^L coincide $\tilde{\mathbb{P}}^*$ -a.s., for each t . Moreover, both sides of the equality in (3.15) above admit right-continuous versions, so it remains to use the density of \mathcal{T} in $[0, T]$ to conclude that $\tilde{\mathbb{P}}^*$ is also weakly admissible. \square

Next, we couple the probability measures $\{\tilde{\mathbb{P}}^{(n)}\}_{n \in \mathbb{N}}$ and $\tilde{\mathbb{P}}^*$ on the same probability space.

Lemma 3.6. *There exists a probability space and on it a sequence $\{(A^{(n)}, L^{(n)}, M^{(n)})\}_{n \in \mathbb{N}}$ of \mathcal{D}^{d+2k} -valued random elements, as well as an \mathcal{D}^{d+2k} -valued random element (A, L, M) such that*

- (1) *the law of $(L^{(n)}, A^{(n)}, M^{(n)})$ is $\tilde{\mathbb{P}}^{(n)}$, and the law of (L, A, M) is $\tilde{\mathbb{P}}^*$, and*
- (2) *For almost all $\omega \in \Omega$, we have*

$$(L_T^{(n)}(\omega), A_T^{(n)}(\omega), M_T^{(n)}(\omega)) \rightarrow (L_T(\omega), A_T(\omega), M_T(\omega))$$

as well as

$$(L_t^{(n)}(\omega), A_t^{(n)}(\omega), M_t^{(n)}(\omega)) \rightarrow (L_t(\omega), A_t(\omega), M_t(\omega))$$

in (Lebesgue) measure in t .

Proof. The first step is use Dudley's extension (see [Dud68], Theorem 3., p. 1569) of the Skorokhod's representation theorem to transform the Meyer-Zheng convergence to an almost-sure convergence in the pseudopath topology. Indeed, the original theorem of Skorokhod cannot be applied directly since the canonical space \mathcal{D}^{d+2k} , together with the pseudopath topology is not Polish. Next, a minimal adjustment of a result of Dellacherie (see Lemma 1., p. 356 in [MZ84]) states that the pseudopath topology and the topology of the convergence in the sum $\lambda + \delta_T$ of the Lebesgue measure λ on $[0, T]$ and the Dirac mass δ_T on $\{T\}$ coincide. \square

On the probability space of Lemma 3.6, we define the sequences

$$N^{(n)} = \int_0^\cdot \nabla h(L_t^{(n)}, A_t^{(n)}) dt, \quad F^{(n)} = \int_0^\cdot f(t) dA_t^{(n)},$$

as well as

$$N = \int_0^\cdot \nabla h(L_t, A_t) dt, \quad F = \int_0^\cdot f(t) dA_t,$$

Using the polynomial-growth assumptions and the \mathbb{L}^p -boundedness of $\{A^{(n)}\}_{n \in \mathbb{N}}$ we see immediately that

$$N^{(n)} \rightarrow N \text{ in } \mathbb{L}^1(\lambda \otimes \mathbb{P}), \text{ and } M_T^{(n)} \xrightarrow{\mathbb{L}^1} M_T.$$

To deal with $\{F^{(n)}\}_{n \in \mathbb{N}}$, we can use an argument completely analogous to that in the last part of the proof of Theorem 2.7 (with K replaced by $[0, T] \setminus \mathcal{T}$). Indeed, together with the \mathbb{L}^p -boundedness of $\{A_T^{(n)}\}_{n \in \mathbb{N}}$, for all $p \geq 1$, it yields that

$$(3.16) \quad F_T^{(n)} \xrightarrow{\mathbb{L}^1} F_T.$$

3.4.5. A passage to a limit in the Pontryagin FBSDE. We define $Y^{(n)} = f + M^{(n)} - N^{(n)}$ so that

$$Y^{(n)} \rightarrow Y = f + M - N \text{ in } \mathbb{L}^1(\lambda \otimes \mathbb{P}).$$

Thus,

$$\mathbb{E}\left[\int_0^T Y_t^- dt\right] = \lim_n \mathbb{E}\left[\int_0^T (Y_t^{(n)})^- dt\right] = \lim_n \mathbb{E}\left[\int_0^T (Y_t^{[n]})^- dt\right] = 0,$$

where the last equality follow directly from equation (3.11) of Proposition 3.3. Consequently, by right continuity,

$$(3.17) \quad Y_t \geq 0 \text{ for all } t \in [0, T].$$

Next, we observe that, by Lemma 3.6 and equation (3.16), we have

$$\mathbb{E}[C(L, A)] = \lim_n \mathbb{E}[C(L^{(n)}, A^{(n)})] = \inf_n \mathbb{E}[C(L^{(n)}, A^{(n)})].$$

Therefore, for each $n \in \mathbb{N}$, we have

$$0 \leq \mathbb{E}[C(L^{(n)}, A^{(n)})] - \mathbb{E}[C(L, A)] =: K_n + I_n,$$

where

$$K_n = \mathbb{E}[C(L^{(n)}, A^{(n)})] - \mathbb{E}[C(L^{(n)}, A)] \text{ and } I_n = \mathbb{E}[C(L^{(n)}, A) - C(L, A)].$$

By convexity of h and g and integration by parts we have

$$\begin{aligned} K_n &\leq \mathbb{E}[\langle \partial C(L^{(n)}, A^{(n)}), A^{(n)} - A \rangle] = \mathbb{E} \left[F_T^{(n)} - F_T \right] + \\ &\quad + \mathbb{E} \left[\int_0^T \nabla h(L_t^{(n)}, A_t^{(n)})(A_t^{(n)} - A_t) dt + \nabla g(L_T^{(n)}, A_T^{(n)})(A_T^{(n)} - A_T) \right] \\ &= \mathbb{E} \left[\int_0^T Y_t^{(n)} dA_t^{(n)} \right] - R_n \end{aligned}$$

where

$$R_n = \mathbb{E} \left[F_T + \int_0^T \nabla h(L_t^{(n)}, A_t^{(n)}) A_t dt + \nabla g(L_T^{(n)}, A_T^{(n)}) A_T \right].$$

By equation (3.10) of Proposition 3.3, we then have

$$K_n \leq -n \mathbb{E} \left[\int_0^T (Y_t^{(n)})^- dt \right] - R_n \leq -R_n.$$

On the other hand, thanks to the growth assumptions, the family $\{C(L^{(n)}, A) - C(L, A)\}_{n \in \mathbb{N}}$ is uniformly integrable. By the continuity of g and h in the l -argument, we have $C(L^{(n)}, A) \rightarrow C(L, A)$ a.s., so $I_n \rightarrow 0$, as $n \rightarrow \infty$. It follows that $\liminf R_n \leq 0$, and, therefore,

$$(3.18) \quad \mathbb{E} \left[F_T + \int_0^T \nabla h(L_t, A_t) A_t dt + \nabla g(L_A, A_T) A_T \right] \leq 0.$$

Next we investigate the martingale properties of the third component process M , in the spirit of the martingale-preservation property of the Meyer-Zheng convergence (see Theorem 11., p. 368 in [MZ84]). On the filtered probability space of the capped problem (i.e., of subsection 3.4), the process $M^{[n]}$ is a martingale, and $A^{[n]}$ is adapted with respect to the augmented filtration generated by L . Thus, we have

$$\mathbb{E} \left[M_t^{(n)} \varphi \left((L_{t_i}^{(n)}, A_{t_i}^{(n)}, M_{t_i}^{(n)})_{1 \leq i \leq k} \right) \right] = \mathbb{E} \left[M_T^{(n)} \varphi \left((L_{t_i}^{(n)}, A_{t_i}^{(n)}, M_{t_i}^{(n)})_{1 \leq i \leq k} \right) \right]$$

for each $k \in \mathbb{N}$, a continuous bounded function $\varphi : \mathbb{R}^{d+2k} \rightarrow \mathbb{R}$ and any choice of $0 \leq t_1 < t_2 < \dots < t_k \leq t$. It follows that, with \mathcal{T} as in Theorem A.4, that

$$(3.19) \quad \mathbb{E}[M_T | \mathcal{F}_t^{L, A, M}] = M_t, \text{ a.s., for } t \in \mathcal{T},$$

and, then, by the right-continuity of the paths of M , that M is an $\mathbb{F}^{(L,A,M)}$ -martingale. The inequality (3.18) implies that after another round of integration by parts - we have

$$(3.20) \quad \int_0^T Y_t dA_t \leq 0, \text{ a.s.}$$

It remains to aggregate the above results to conclude that the (law) of the triplet (L, A, Y) is a weak solution of the Pontryagin FBSDE (Definition 2.10). Part (1) is exactly the content of Proposition 3.5, while part (2) follows from (3.17) and (3.20). Finally (3) is simply a restatement of the martingale property of the process M , established after (3.19) above. Theorem 2.12, part (2) now allows us to conclude that the law of the pair (L, A) is a solution to the monotone follower problem.

APPENDIX A. AUXILIARY RESULTS

In this appendix we gather several results that are used in the body of the paper. They either admit hard-to-locate standard proofs, or are minimal extensions of the known results; we state them here, and supply proofs, for completeness sake.

A.1. Coupling of weakly admissible controls. We start simple coupling lemma based on a standard use of regular conditional probabilities. It is used in proofs of Theorem 2.7 and Theorem 2.21 above.

Lemma A.1 (Coupling). *For $d, k, l \in \mathbb{N}$, let $\mathbb{P} \in \mathfrak{P}^{d+k}(L, Q)$ and $\mathbb{P}' \in \mathfrak{P}^{d+l}(L', R')$ be such that $\mathbb{P}_L = \mathbb{P}'_{L'}$. Then, there exists a probability measure $\bar{\mathbb{P}} \in \mathfrak{P}^{d+k+l}(\bar{L}, \bar{Q}, \bar{R})$, denoted by $\mathbb{P} \otimes_L \mathbb{P}'$ such that*

- (1) $\bar{\mathbb{P}}_{\bar{L}, \bar{Q}} = \mathbb{P}_{L, Q}$,
- (2) $\bar{\mathbb{P}}_{\bar{L}, \bar{R}} = \mathbb{P}'_{L', R'}$, and
- (3) \bar{Q} and \bar{R} are $\bar{\mathbb{P}}$ -conditionally independent, given \bar{L} .

Proof. The space $\mathcal{D}^d(L, Q)$ is a Borel space, so there exists a regular conditional distribution (r.c.d.)

$$\mu : \mathcal{D}^d(L, Q) \times \mathcal{B}(\mathcal{D}^k) \rightarrow [0, 1], \quad \mu(x, B) = \mathbb{P}[Q \in B | L = x],$$

for Q , given L under \mathbb{P} . Similarly, let $\mu' : \mathcal{D}^d(L', R') \times \mathcal{B}(\mathcal{D}^l) \rightarrow [0, 1]$ denote the \mathbb{P}' -r.c.d. of R given L' and let ρ denote the the product kernel $\rho : \mathcal{D}^d \times \mathcal{B}(\mathcal{D}^{k+l}) \rightarrow [0, 1]$, given by

$$\rho(x, B) = (\mu(x, \cdot) \otimes \nu(x, \cdot))(B), \text{ for } x \in \mathcal{D}^d \text{ and } B \in \mathcal{B}(\mathcal{D}^{k+l}).$$

We define $\bar{\mathbb{P}}$ as the (Ionescu-Tulcea-type) product $\mathbb{P}_L \otimes \rho$ of the measure \mathbb{P}_L and the kernel ρ , i.e., the probability measure given by

$$\bar{\mathbb{P}}[C] = \int_{x \in \mathcal{D}^d} \int_{(q, r) \in \mathcal{D}^{k+l}} \mathbf{1}_C(x, q, r) \rho(x, dq, dr) \mathbb{P}_L(dx),$$

for $C \in \mathcal{B}(\mathcal{D}^{d+k+l})$. The reader will readily check that so defined, $\bar{\mathbb{P}} = \mathbb{P} \otimes_L \mathbb{P}'$ satisfies all three conditions in the statement. \square

An immediate application of Lemma A.1 is the following

Lemma A.2. *Let $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$ be a sequence in \mathcal{A} . Then, there exists a probability space and, on it, càdlàg processes $\{L_t\}_{t \in [0, T]}$, $\{A_t^{(n)}\}_{t \in [0, T]}$, $n \in \mathbb{N}$, such that the joint law of $(L, A^{(n)})$ is \mathbb{P}_n , for each $n \in \mathbb{N}$, and $\{A^{(n)}\}_{n \in \mathbb{N}}$ are independent, conditionally on L .*

Proof. We can think of the required sequence $L, A^{(1)}, A^{(2)}, \dots$ as a stochastic process with values in \mathcal{D}^k (and \mathcal{D}^d for its first component). Using the information on the joint distributions and the requirement of conditional independence from the statement, we can apply Lemma A.1 repeatedly to construct its (consistent) family of finite-dimensional distributions. The target spaces \mathcal{D}^d and \mathcal{D}^k are Polish, so the sought-for probability space $(\Omega, \mathcal{F}, \mathbb{P})$ can now be constructed by using Kolmogorov's extension theorem. \square

A.2. An \mathbb{L}^p estimate.

Lemma A.3. *Given $p \geq 1$, suppose that $X \in \mathbb{L}_+^1$ and $Y \in \mathbb{L}_+^p$ satisfy*

$$(A.1) \quad \mathbb{E}[(X - r)^+] \leq \mathbb{E}[Y 1_{\{X > r\}}], \quad \text{for all } r \geq 0.$$

Then, $X \in \mathbb{L}^p$ and $\|X\|_p \leq p\|Y\|_p$.

Proof. The conclusion clearly holds for $p = 1$: it suffices to substitute $r = 0$ into (A.1). For $p > 1$, multiplying both sides of (A.1) by $(p - 1)r^{p-2}$ and integrating in r over $[0, M]$, for $M > 0$, yields

$$\begin{aligned} \mathbb{E}[Y(X \wedge M)^{p-1}] &= \mathbb{E}\left[Y \int_0^M (p-1)r^{p-2} 1_{\{X > r\}} dr\right] \\ &\geq \mathbb{E}\left[\int_0^M (p-1)r^{p-2} (X-r)^+ dr\right] \geq \frac{1}{p} \mathbb{E}[(X \wedge M)^p]. \end{aligned}$$

It remains to apply Hölder's inequality to obtain

$$\frac{1}{p} \mathbb{E}[(X \wedge M)^p] \leq \mathbb{E}[Y(X \wedge M)^{p-1}] \leq \|Y\|_p \|X \wedge M\|_p^{p-1},$$

which, after dividing both sides by $\|X \wedge M\|_p^{p-1}$, and letting $M \rightarrow \infty$, completes the proof. \square

A.3. The pseudopath topology. The topology τ_{pp} we consider on \mathcal{D}^N is a following minimal modification of the pseudopath topology introduced in [MZ84].

A path $x \in \mathcal{D}^N$ can be identified with its **pseudopath**, i.e., a finite measure on the product $[0, T] \times \mathbb{R}^N$, obtained as a push-forward of the “reinforced” Lebesgue measure $\text{Leb} + \delta_{\{T\}}$ on $[0, T]$, where $\delta_{\{T\}}$ denotes the Dirac mass at $\{T\}$, via the map

$$[0, T] \ni t \mapsto (t, x(t)) \in [0, T] \times \mathbb{R}^N.$$

With such an identification, the trace of the topology of weak convergence of measures is induced on \mathcal{D}^N ; we call it the **pseudopath topology** and denote by τ_{pp} . It is shown in [MZ84, Lemma 1, p. 365] - we modify this result (and all others) minimally to fit our setting - that the pseudopath topology is metrizable and that, for a sequence $\{x_n\}_{n \in \mathbb{N}}$ in \mathcal{D} , we have $x_n \xrightarrow{\text{PP}} x \in \mathcal{D}$, where $\xrightarrow{\text{PP}}$ denotes the convergence in the pseudopath topology, if and only if

$$(A.2) \quad x_n(T) \rightarrow x(T) \text{ and } \int_0^T b(s, x_n(s)) ds \rightarrow \int_0^T b(s, x(s)) ds,$$

for all continuous and bounded functions $b : [0, T] \times \mathbb{R}^N \rightarrow \mathbb{R}$. Finally, we mention a result due to Dellacherie (see [MZ84], Lemma 1, p. 356) which simply states that the convergence in the pseudopath topology and the convergence in the measure $\lambda + \delta_{\{T\}}$ coincide.

A.4. The Meyer-Zheng convergence. Using the pseudopath topology τ_{pp} on \mathcal{D}^N , one can define the **Meyer-Zheng topology** on \mathfrak{P}^N as the topology of weak convergence of probability measures on the topological space $(\mathcal{D}^N, \tau_{pp})$. Like the pseudopath topology τ_{pp} on \mathcal{D}^N , the Meyer-Zheng topology on \mathfrak{P} is metrizable, but not necessarily Polish (see p. 372 in [MZ84]); the convergence in the Meyer-Zheng topology is denoted by $\xrightarrow{\text{MZ}}$. As shown in [MZ84], the Borel σ -algebra generated by the pseudopath topology τ_{pp} coincides with the canonical σ -algebra on \mathcal{D}^N , i.e., the one induced by the coordinate maps or, equivalently, by the Skorokhod topology. Moreover, the set of all pseudopaths, denoted by Ψ , under τ_{pp} is Polish.

We note the following (minimal extension) of a useful consequence of the Meyer-Zheng convergence (see [MZ84], Theorem 5., p. 365):

Theorem A.4 (Meyer and Zheng, 1984). *Let $\{\mathbb{P}^n\}_{n \in \mathbb{N}}$ be a sequence of probability measures on \mathcal{D}^N such that $\mathbb{P}^n \rightarrow \mathbb{P}$ in the Meyer-Zheng sense. Then there exists a subset $\mathcal{T} \subseteq [0, T]$ of full Lebesgue measure, containing T , such that the \mathbb{P}^n -finite-dimensional distributions with indices in \mathcal{T} of the coordinate process converge to the corresponding finite-dimensional distributions under \mathbb{P} , perhaps after a passage to a subsequence.*

A.5. A criterion for compactness. One of the reasons the Meyer-Zheng topology proved to be quite useful in probability theory and optimal stochastic control is a simple characterization of compactness it affords. Unlike the Skorokhod topology, where compactness needs a stronger form of equicontinuity, the subsets of \mathfrak{P}^N are Meyer-Zheng-compact as soon as they are suitably bounded. The following result is a compilation of two statements in [MZ84], namely Theorem 4., p. 360, and Theorem 5., p. 365, minimally adapted to fit our setting. We remind the reader that an adapted stochastic process X , defined on a filtered measurable space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0, T]})$ is said to be a **quasi-martingale** under the probability measure \mathbb{P} if $X_t \in \mathbb{L}^1(\mathbb{P})$, for all $t \in [0, T]$ and $\text{Var}^\mathbb{P}[X] < \infty$, where

$$(A.3) \quad \text{Var}^\mathbb{P}[X] = \sup \sum_{j=1}^m \mathbb{E}^\mathbb{P} \left[\left| \mathbb{E}^\mathbb{P} [X_{t_j} - X_{t_{j-1}} | \mathcal{F}_{t_j}] \right| \right] + \mathbb{E}^\mathbb{P} [|X_T|],$$

and the supremum is taken over all partitions $0 = t_0 < \dots < t_m = T$, $m \in \mathbb{N}$, of $[0, T]$.

Theorem A.5 (Meyer and Zheng, 1984). *Let $\{\mathbb{P}^n\}_{n \in \mathbb{N}}$ be a sequence of probability measures on \mathcal{D}^N (equipped with the filtration generated by the coordinate maps) with the property that each coordinate process $\{X_t^i\}_{t \in [0, T]}$, $i = 1, \dots, N$, is a \mathbb{P}^n -quasimartingale for each $n \in \mathbb{N}$ and*

$$\sup_{n \in \mathbb{N}} \text{Var}^{\mathbb{P}^n}[X^i] < \infty, \text{ for all } i = 1, \dots, N.$$

Then, there exists a subsequence $\{\mathbb{P}_{n_k}\}_{k \in \mathbb{N}}$ of $\{\mathbb{P}^n\}_{n \in \mathbb{N}}$ and $\mathbb{P} \in \mathfrak{P}$ such that $\mathbb{P}_{n_k} \xrightarrow{\text{MZ}} \mathbb{P}$ in the Meyer-Zheng topology.

Remark A.6. The condition $\sup_n \text{Var}^{\mathbb{P}_n}[X^i] < \infty$ is easy to check if X^i is a \mathbb{P}_n martingale, for each $n \in \mathbb{N}$. Indeed, in that case $\text{Var}^{\mathbb{P}_n}[X^i] = \mathbb{E}^{\mathbb{P}_n}[|X_T^i|]$, with its boundedness being equivalent to uniform \mathbb{L}^1 -boundedness of the process X^i under all $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$.

Similarly, if X^i happens to be a process of finite variation, $\text{Var}^{\mathbb{P}_n}[X^i]$ is bounded from above by a (constant multiple) of the expected total variation of X^i . In particular, if X^i is nonnegative and nondecreasing under all \mathbb{P}_n , the condition we are looking for is exactly the same as in the martingale case: $\sup_n \mathbb{E}^{\mathbb{P}_n}[|X_T^i|] < \infty$.

REFERENCES

- [AM03] ANTONELLI, F. AND MA, J. Weak solutions of forward-backward SDE's. *Stochastic Anal. Appl.*, 21(3):493–514, 2003.
- [Ban05] BANK, P. Optimal control under a dynamic fuel constraint. *SIAM J. Control Optim.*, 44(4):1529–1541 (electronic), 2005.
- [BC67] BATHER, J. AND CHERNOFF, H. Sequential decisions in the control of a spaceship. In *Proc. Fifth Berkeley Sympos. Mathematical Statistics and Probability (Berkeley, Calif., 1965/66), Vol. III: Physical Sciences*, pages 181–207. Univ. California Press, Berkeley, Calif., 1967.
- [BC09] BAIN, A. AND CRISAN, D. *Fundamentals of stochastic filtering*, volume 60 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2009.
- [BR01] BANK, P. AND RIEDEL, F. Optimal consumption choice with intertemporal substitution. *Ann. Appl. Probab.*, 11(3):750–788, 2001.
- [BR06] BUDHIRAJA, A. AND ROSS, K. Existence of optimal controls for singular control problems with state constraints. *Ann. Appl. Probab.*, 16(4):2235–2255, 2006.
- [BY78] BRÉMAUD, P. AND YOR, M. Changes of filtrations and of probability measures. *Z. Wahrsch. Verw. Gebiete*, 45(4):269–295, 1978.
- [CH94] CADENILLAS, A. AND HAUSSMANN, U. G. The stochastic maximum principle for a singular control problem. *Stochastics Stochastics Rep.*, 49(3-4):211–237, 1994. ISSN 1045-1129.
- [CM96] CVITANIĆ, J. AND MA, J. Hedging options for a large investor and forward-backward SDE's. *Ann. Appl. Probab.*, 6(2):370–398, 1996.
- [Dud68] DUDLEY, R. M. Distances of probability measures and random variables. *Ann. Math. Statist.*, 39:1563–1572, 1968.
- [GT08] GUO, X. AND TOMECEK, P. Connections between singular control and optimal switching. *SIAM J. Control Optim.*, 47(1):421–443, 2008.
- [HS95] HAUSSMANN, U. G. AND SUO, W. Singular optimal stochastic controls. I. Existence. *SIAM J. Control Optim.*, 33(3):916–936, 1995.
- [Kab99] KABANOV, Y. Hedging and liquidation under transaction costs in currency markets. *Finance and Stochastics*, 3(2):237–248, 1999.
- [Kal02] KALLENBERG, O. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.
- [KS84] KARATZAS, I. AND SHREVE, S. E. Connections between optimal stopping and singular stochastic control. I. Monotone follower problems. *SIAM J. Control Optim.*, 22(6):856–877, 1984.
- [MC01] MA, J. AND CVITANIĆ, J. Reflected forward-backward SDEs and obstacle problems with boundary conditions. *J. Appl. Math. Stochastic Anal.*, 14(2):113–138, 2001.
- [MPY94] MA, J., PROTTER, P., AND YONG, J. M. Solving forward-backward stochastic differential equations explicitly—a four step scheme. *Probab. Theory Related Fields*, 98(3):339–359, 1994.
- [MY99] MA, J. AND YONG, J. *Forward-backward stochastic differential equations and their applications*, volume 1702 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1999.
- [MZ84] MEYER, P.-A. AND ZHENG, W. A. Tightness criteria for laws of semimartingales. *Ann. Inst. H. Poincaré Probab. Statist.*, 20(4):353–372, 1984.

- [MZ11] MA, J. AND ZHANG, J. On weak solutions of forward-backward SDEs. *Probab. Theory Related Fields*, 151(3-4):475–507, 2011.
- [Pra99] PRATELLI, M. An alternative proof of a theorem of Aldous concerning convergence in distribution for martingales. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 334–338. Springer, Berlin, 1999.
- [RS11] RIEDEL, F. AND SU, X. On irreversible investment. *Finance Stoch.*, 15(4):607–633, 2011.
- [Sch86] SCHWARTZ, M. New proofs of a theorem of Komlós. *Acta Math. Hung.*, 47:181–185, 1986.
- [Sch12] SCHNURR, A. On the semimartingale nature of Feller processes with killing. *Stochastic Process. Appl.*, 122(7):2758–2780, 2012.
- [Ste12] STEG, J.-H. Irreversible investment in oligopoly. *Finance Stoch.*, 16(2):207–224, 2012.

JIEXIAN LI, DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF TEXAS AT AUSTIN

E-mail address: jxli@math.utexas.edu

GORDAN ŽITKOVIĆ, DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF TEXAS AT AUSTIN

E-mail address: gordanz@math.utexas.edu